View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Waseda University Repository

修 士 論 文 概 要 書

2010 年　2 月提出

専攻名（専門分野）	情報理工学専攻	氏 名	木佐森　幸太	指　導 教　員	後藤　滋樹　　印
研究指導名	後藤　滋樹	学籍番号	CD 5108B034-7		
研　究 題　目	TCP フィンガープリントによる悪意のある通信の分析				

概要: カーネルマルウェアは独自のネットワークドライバを実装し、カーネルモードの通信を行うことで監視ツールからの隠匿を試みる。これらのネットワークドライバは独自の実装であるため、通信の特徴を分析することで既存OS による通信から区別することが可能である。本研究ではハニーポットの通信データを分析対象とし，既存OS とは異なる特徴を有する通信を抽出し、その挙動を分析してカーネルマルウェアの可能性があるホストを検出する手法を提案する。また、実運用網の通信データにそれを適用し、その有効性を実証する。

1 カーネルマルウェアとTCP fingerprinting

1.1 カーネルマルウェア

カーネルマルウェアとは、CPUの最も権限レベルが高いモード（カーネルモード）で動作するマルウェアである。これらのマルウェアは独自のネットワークドライバを用いてカーネルモードでのネットワーク通信を行うが、このネットワークドライバはOS由来のTCP/IPとは異なる実装であるため、TCP/IPヘッダの特徴を分析することで既存OS由来の通信と区別することができる。

1.2 TCP fingerprinting

TCP/IPの仕様はRFCで定義されているが、OSごとにその実装は異なっている。そのため、通信データを分析することで対象システムのOSを推定することができる。このような技術をfingerprintingという。本研究では、受信した通信データを分析することでOSを分析するPassive fingerprintingを行う代表的なツールであるp0fを用いる。

1.3 p0f

p0fにはいくつかのモードがあるが、本研究ではSYNパケットのみを分析対象とするモードを用いる。このモードではSYNパケットのTCP/IPヘッダにおける各種の値（ウィンドウサイズ、TTLの初期値、DFビット、TCPオプションなど）をシグネチャとして持ち、OSを推定する。

1.4 提案手法

本研究では、まずハニーポットの通信データにp0fを適用して既存OSによる通信ではないと判別された通信を抽出し、シグネチャとして集約する。抽出されたシグネチャをベースとして通信データの分析を行い、独自のネットワークドライバを用いたカーネルマルウェアによる悪意のある通信であることを検証する。次に、抽出したシグネチャを外部の実運用ネットワーク通信データに適用し、その結果を分析することで提案手法の有効性を実証する。

2 CCC DATAsetの分析

CCC DATAsetとは、サイバークリーンセンター（CCC）で運用しているハニーポットのデータから作成されたマルウェア研究用データセットである。本研究ではハニーポット2台の通信を2日間フルキャプチャしたものを用いる。収集期間は2008年4月28日、29日と2009年3月13日、14日である。

通信データにp0fを適用したところ、既存のOSではないと判定された通信が多数検出された。これらについて、TTLを2の累乗の値に切り上げて初期値と推定することで、44のシグネチャに集約した。以降、MWSシグネチャと総称する。また、個々のシグネチャの第一フィールド（ウィンドウサイズ）によって名称を付けた。ウィンドウサイズが65535のシグネチャであればMWS 65535_1、MWS 65535_2…となる。

分析の結果、SYNパケット、送信元IPアドレスの約半数がMWSシグネチャによるものであった。個々のシグネチャごとにSYNパケット数、送信元IPアドレス数を集計した結果、やはりMWSシグネチャが上位に現れた。

出現頻度の高いいくつかのMWSシグネチャについて送信先ポート番号の集計を行ったところ、135番、139番、445番、1433番、2967番といった著名な脆弱性があるポート番号への通信の比率が高いことが分かった。悪意のある通信が行われている可能性が高いと考えられる。

また、これらのシグネチャを有するホストの攻撃パターン分析も行った。脆弱性のあるポートへのスキャンや攻撃、マルウェアのダウンロードなどが行われていることがわかった。

3 他のネットワークのデータへの応用

3.1 MWSシグネチャの拡張

CCC DATAsetでは、すべての通信のDFビットが0であった。DFビットはルータやファイアウォールによって削除されることがあり、CCC DATAsetの収集環境でも経路上で削除された可能性がある。そのため、MWSシグネチャのDFビットを1 としたMWS+DFシグネチャを作成した。

また、MWSシグネチャの中には最大セグメントサイズ（MSS）オプションの値のみが異なるグループが6群存在した。MSSの値は通信環境によって左右されるものであるため、これらのグループについてはDFビットを1としたうえでMSSの値をワイルドカードとし、6種のMWS_Genシグネチャを作成した。

3.2 早稲田大学の通信データ

早稲田大学の対外接続回線のTCP SYNパケットを収集したものである。収集期間は2009年12月25日から12月31日までの1週間である。各種MWSシグネチャについて、SYNパケット数と送信元IP数が全体に占める割合を表1

に示す。

表1: 早稲田大学の通信データ統計

	MWS	MWS+DF	MWS_Gen
SYNパケット数	5.140%	0.904%	2.569%
送信元IP数	0.007%	0.770%	4.656%

MWSシグネチャによるSYNパケット数が送信元IP数に比べて多い理由は、少数のIPからMWS 16384_1シグネチャによる多量のSYNパケット送信があったためである。それ以外にはMWS_GenシグネチャによるSYNパケット数が多かった。

上位4種のMWS_Genシグネチャについて送信先ポート番号を分析したところ、135番、139番、445 番、2967番、1433番のほかにHTTP/HTTPS、SMTP、bittorentで用いられるポート番号などが上位に入っていることが分かった。

3.3 企業におけるSMTPデータ

ある企業の電子メールサーバに接続したネットワークセグメントで収集したTCPヘッダデータであり、この回線で観測可能な通信はSMTPのみである。収集期間は2009年3月1日から3月31日までの1カ月間である。各種MWSシグネチャについて、SYNパケット数と送信元IP数が全体に占める割合を表2に示す。

表2: 企業におけるSMTP データ統計

	MWS	MWS+DF	MWS_Gen
SYNパケット数	0.004%	0.794%	2.107%
送信元IP数	0.004%	0.240%	3.500%

シグネチャごとのSYNパケット送信数の上位には、MWS_GenシグネチャやMWS+DF 8192_1シグネチャが現れたが、MWS 16384_1シグネチャによるSYNパケット送信は観測されなかった。また、MWSシグネチャによるSYNパケット送信があったIPアドレスについて、発信したメールの内容の判別を行ったところ、数は少ないもののすべてのメールがスパムメールと判別された。マルウェアの構成によってはスパム送信モジュールを搭載するものも存在すると推定できる。

3.4 MAWIデータセット

MAWIデータセットとは、WIDE Projectによって行われているインターネット定点観測にて収集されているものである。今回は、データセットのうち太平洋を横断するネットワーク回線において、毎日15分間取得されているフルキャプチャデータを用いた。使用した期間は2006年11月から2009年11月の37カ月間である。月ごとに集計を行い、SYNパケット数、送信元IP数について全体に占める割合の推移について分析した。

SYNパケット数については、すべての月においてMWSシグネチャによるSYNパケットのおよそ99% 以上がMWS 16384_1シグネチャによるものであった。MWS 16384_1シグネチャ以外のMWSシグネチャによるSYNパケットは、存在しているもののほぼ0%である。また、月による変動が激しいが、多くの月においてMWS+DFシグネチャ、MWS_GenシグネチャによるSYNパケットが全体のおよそ1.5%以上存在している。

送信元IPについて、MWSシグネチャの送信元IPはどの月でも0.1%未満と非常に少ないが、MWS+DFシグネチャ、MWS_Genシグネチャの送信元IPを合計すると、2007年7月を除いては2%を上回っており、SYNパケット数の割合の推移よりも安定して推移している。2007年7月以降は増加傾向にあり、2009年は4〜5%で推移している。

3.5 MWS 16384_1シグネチャ

早稲田大学、MAWI双方のデータでMWS 16384_1シグネチャによる大量のSYNパケットが観測された。共通するのは、少数の送信元IPから多量のSYNパケットが送信されていることである。また、送信先ポート番号については、既出の著名な脆弱性のあるポートのほか、Oracleデータベースや MySQLで用いられるポート番号が上位に現れている。

最も特徴的なのは、大多数のSYNパケットの送信元ポートが6000番であることである。その割合は95%前後に達する。いくつかの送信元IPアドレスについては、MWS 16384_1シグネチャと既存OSのシグネチャの両方によるSYNパケット送信が見られた。

3.6 考察

MWSシグネチャの拡張を行った結果、それらのシグネチャによる通信が多数観測された。この結果より、MWSシグネチャのDFビットは本来1であり、CCC DATAsetの収集環境によって0になっている可能性が高いと考えられる。また、MWS_Genシグネチャについても送信先ポート番号の上位に著名な脆弱性のあるポートが存在することから、やはり悪意のある通信が行われていることが疑われる。

それぞれのデータセットにおけるシグネチャの出方には差異があった。これは通信環境の違いに起因するものと考えられる。収集するポート番号の違い、データ収集のポイントの違い、ネットワークアドレスの違いなどである。

4 まとめ

本論文では、カーネルマルウェアの可能性がある通信をTCPフィンガープリントによって識別する手法を提案し、実ネットワーク上の通信データの分析を通して提案手法の有効性を検討した。ハニーポットに対する通信だけでなく、実ネットワーク上の通信データでも、作成したシグネチャによる悪意のある通信が行われている可能性が高いことを示すことができた。

今後の課題として、よりパケットの内容に踏み込んだ分析を行う必要があること、シグネチャの送信元ホストを特定して分析をすること、シグネチャを随時アップデートすること等が挙げられる。

参考文献

[1] Michal Zalewski, “the new p0f: 2.0.8”, http://lcamtuf.coredump.cx/p0f.shtml, 2006.
[2] 木佐森幸太, 下田晃弘, 森達哉, 後藤滋樹, “TCPフィンガープリントによる悪意のある通信の分析”, マルウェア対策研究人材育成ワークショップ2009 (MWS2009), A6-2, pp553-558, 2009.